AFHRL-TP-89-74

AIR FORCE

HUMAN RESOURCES

LABORATORY

ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB): ITEM, OVERLENGTH AND OPERATIONAL LENGTH DEVELOPMENT OF FORMS 18 AND 19

Linda T. Curran

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Pamla Palmer

Performance Metrics, Inc.
5825 Callaghan Road, Suite 225
San Antonio, Texas 78228

February 1990
Final Technical Paper for Period June 1984 - July 1989

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

This paper has been reviewed and is approved for publication.

JAMES A. EARLES
Contract Monitor

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

DANIEL L. LEIGHTON, Colonel, USAF
Chief, Manpower and Personnel Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
February 1990

3. REPORT TYPE AND DATES COVERED
Final - June 1984 to July 1989

4. TITLE AND SUBTITLE
Armed Services Vocational Aptitude Battery (ASVAB): Item, Overlength and Operational Length Development of Forms 18 and 19

5. FUNDING NUMBERS
C - F41689-84-D-0002
PE - 62703F
PR - 7719
TA - 18
WU - 40

6. AUTHOR(S)
Linda T. Curran
Pamla Palmer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Performance Metrics, Inc.
5825 Callaghan Road, Suite 225
San Antonio, Texas 78228-1178

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFHRL-TP-89-74

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The Armed Services Vocational Aptitude Battery (ASVAB) is the aptitude test used by each of the armed services in making enlisted selection and classification decisions. The objective of this effort was to develop four new forms of the ASVAB for the Department of Defense (DoD) Student Testing Program (STP). These new forms are to be designated 18a, 18b, 19a, and 19b and must be statistically and content parallel to a reference test, ASVAB Form 8a. This paper documents the first two phases of the developmental process. In Phase I, experimental items paralleling the taxonomy of the reference test, Form 8a, were written and administered to Air Force recruits at Lackland Air Force Base, Texas. These experimental items were administered in an equivalent groups design with the reference test items. Item statistics from this administration were calculated and used to match experimental with reference test items in order to develop overlength forms. In Phase II, the overlength forms were administered to recruits from each of the services and again item statistics and item content were used to match experimental with reference form test items in order to create operational length forms. Trial equatings were performed to determine if scores on the new subtests correspond to each other and to the scores on like-named subtests on Form 8a. These equatings and a comparison of the content, item and test statistics between the new forms and ASVAB Forms 8a indicated that the new forms were content and statistically parallel to one another and to the reference tests.

14. SUBJECT TERMS
Armed Services Vocational Aptitude Battery, equipercentile equating, parallel test, test development

15. NUMBER OF PAGES
36

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

AFHRL Technical Paper 89-74 January 1990

ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB): ITEM, OVERLENGTH AND OPERATIONAL LENGTH DEVELOPMENT OF FORMS 18 AND 19

Linda T. Curran

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Pamla Palmer

Performance Metrics, Inc.
5825 Callaghan Road, Suite 225
San Antonio, Texas 78228

Reviewed by

Malcolm James Ree
Force Acquisition Branch

Submitted for publication by

Lonnie D. Valentine, Jr.
Chief, Force Acquisition Branch

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

The Armed Services Vocational Aptitude Battery (ASVAB) is a multiple aptitude test battery composed of 10 subtests that is used by each of the armed services to select and classify enlisted personnel. The purpose of this effort was to develop four new ASVAB forms for administration in the high schools as part of the Student Testing Program (STP). The STP serves to provide test results that can be used to identify individuals who are interested in the military and who meet enlistment qualification standards and serves as a counseling tool to aid students in pursuing careers.

The development of new ASVAB forms typically is accomplished in four phases. This paper documents the first two phases of this process for ASVAB Forms 18 and 19. Phase I involved developing and administering a large pool of items to military recruits from which overlength ASVAB subtests were developed. Further culling of items was accomplished in Phase II when operational length forms were developed to be content and statistically parallel to one another and to a reference test, ASVAB Form 8a. In order to examine the comparability of the operational length subtests and the like-named subtests on the reference form, a detailed analysis of the statistical equivalence of the experimental and reference subtests was accomplished. A comparison of item and test statistics between the reference and new forms indicated that the objective of developing new forms parallel to one another and the reference form was met. Equatings were performed for the newly constructed forms.

PREFACE

This paper documents the efforts conducted under two projects in support of Research, Development and Validation of Selection and Classification Procedures (Contract F41689-84-D-0002). These research and development (R & D) efforts were conducted under the Development of the Armed Services Vocational Aptitude Battery Form D (Items) and the Development of the Armed Services Vocational Aptitude Battery Forms D and E (Overlength and Operational Tests) by Performance Metrics, Inc., San Antonio, Texas, under subcontract to Universal Energy Systems Inc., Dayton, Ohio.

Special appreciation is expressed to Mr Carl S. Haywood and Mr William M. Lee for their programming and documentation contributions and to Drs Malcolm James Ree, Air Force Human Resources Laboratory, Benjamin Fairbank and C. Wayne Shore, Performance Metrics, Inc., for their technical insights. In addition, suggestions by Mr James Earles of the Manpower and Personnel Division of the Air Force Human Reources Laboratory were most helpful. The contributions of these individuals were essential to these projects.

TABLE OF CONTENTS

LIST OF TABLES

ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB): ITEM, OVERLENGTH AND OPERATIONAL LENGTH DEVELOPMENT OF FORMS 18 AND 19

I. INTRODUCTION

The Armed Services Vocational Aptitude Battery (ASVAB) is the selection and classification instrument used for enlistment qualification and job placement in the Army, Navy, Air Force, Marine Corps, and Coast Guard. The Armed Services periodically require the development of new ASVAB forms due to test compromise or test obsolescence resulting from the changing needs and requirements of the various armed services. ASVAB forms are developed for two programs. In the enlistment or production program, the ASVAB is administered annually to about 1 million applicants in military entrance processing stations (MEPS), mobile examining team sites (METs), and Office of Personnel Management (OPM) sites. In the Department of Defense (DoD) Student Testing Program (STP), the ASVAB is administered annually to approximately 1.3 million high school students in over 14,000 high schools (DoD, 1984). The DoD STP serves two purposes: to provide test results that are useful for educational and career counseling and to provide the military services with lists that identify individuals interested in the military and who meet enlistment qualification standards. The purpose of this effort was to develop four new ASVAB forms for administration in the high schools.

ASVAB Content

The ASVAB is a multiple aptitude battery that consists of 10 subtests (Table 1) and measures verbal, quantitative, mechanical, and speeded aptitudes (Ree, Mullins, Mathews, & Massey, 1982). Two of the subtests, Numerical Operations and Coding Speed, are highly speeded. The other eight subtests are power subtests that allow enough time for a majority of students to complete them. Two of the subtests, Paragraph Comprehension and Word Knowledge, are summed to form a verbal composite (called VE) that is sometimes used as if it were an eleventh subtest.

Scores from the subtests are aggregated to form composite measures that are used by the different services and the STP. At the time of this study, four of the subtests, Arithmetic Reasoning, Paragraph Comprehension, Word Knowledge, and Numerical Operations comprised the Armed Forces Qualification Test (AFQT) which is used to report to DoD and Congress as one measure of the quality of the enlisted force. With the implementation of ASVAB Forms 15, 16, and 17 in January 1989, the AFQT was changed to the sum of Arithmetic Reasoning, Paragraph Comprehension, Word Knowledge, and Mathematics Knowledge. In the STP, the AFQT is used to provide recruiters with leads of individuals who potentially qualify for enlistment in the armed services.

Two sets of composites, the Academic and Occupational composites, are reported to the students for counseling purposes. Table 2 shows the subtests in these composites. The Academic composites provide traditional measures related to educational experience and are useful for predicting performance in school courses. The Occupational composites are more complex and are empirically derived from military validity studies and can be used to estimate how well students would perform in different types of military training and occupations.

Table 1. Subtests of the ASVAB[a]

Subtest	Code	Number of items	Time in minutes	Contents
General Science	GS	25	11	Physical, life and earth sciences
Arithmetic Reasoning	AR	30	36	Arithmetic word problems
Word Knowledge	WK	35	11	Meaning of selected words
Paragraph Comprehension	PC	15	13	Understanding of written material from brief paragraphs
Numerical Operations	NO	50	3	Speeded numerical calculations
Coding Speed	CS	84	7	Speeded use of a key that matches words and numbers
Auto and Shop Information	AS	25	11	Automobile, tools and shop terminology and practices
Mathematics Knowledge	MK	25	24	Application of learned mathematical principles
Mechanical Comprehension	MC	25	19	Use of mechanical and physical principles
Electronics Information	EI	20	9	Simple electrical or electronics knowledge
TOTAL		334	144	

[a] *VE = WK + PC and is treated as if it were an eleventh subtest.*

Table 2. Definitions of Selected ASVAB Composites

Composite	Subtest composition
Academic Composites	
Verbal	WK + PC + GS
Math	AR + MK
Academic Ability	WK + PC + AR
Occupational Composites	
Mechanical and Crafts	AR + AS + MC + EI
Business and Clerical	VE + CS + MK
Electronics and Electrical	GS + AR + MK + EI
Health/Social/Technology	AR + VE + MC

ASVAB Development Process

When forms of the ASVAB are developed, the new forms must be content and statistically parallel to one another and at least content parallel to a reference test, ASVAB Form 8a (Ree, Mathews, Mullins, & Massey, 1982). Tests are statistically parallel if the tests have equivalent raw score means, variances, and reliabilities. Parallel forms are required in order to equate the new forms to the reference form. Equating enables the armed services to compare the distributions of ability of current applicants to previous applicants and to provide a consistent meaning for the cutting scores used in selection and classification of enlisted personnel.

ASVAB Form 8a was established as the reference test when it was administered by the National Opinion Research Center (NORC) of the University of Chicago to about 12,000 youths from July through October 1980 (Office of the Assistant Secretary of Defense, 1982). The sample was weighted to represent American youth, ages 16 to 23 years of age. A subsample of 18- through 23-year-old males and females formed the reference population for the ASVAB. This subsample is referred to as the 1980 Youth Population and was used to establish the 1980 score scale or metric.

New ASVAB forms are developed through an iterative process that successively culls candidate test items in order to create forms that are parallel to one another and ASVAB Form 8a. This process is accomplished in four phases. Note that in all the phases the candidate or experimental items are administered in conjunction with the Form 8a test items to provide a means of determining which experimental items should be culled and to establish the ultimate relationship between the new and the reference forms. In addition, across phases samples representing successively closer approximations to the target population are used. Phase I involves developing and administering a large pool of items from which overlength ASVAB forms are subsequently developed. Further culling of items is accomplished in Phase II when operational length forms are developed. In Phase III, the operational length forms and the reference test are administered to develop conversion tables which place the new forms on the scale of the reference test. These tables are used in the Initial Operational Test and Evaluation (IOT&E) of the new forms in Phase IV. The new forms and the reference test are administered in an operational setting to develop the final conversion tables. This paper documents the first two phases of this developmental process for ASVAB Forms 18 and 19.

II. PHASE I - DEVELOPMENT OF ITEMS AND OVERLENGTH FORMS

The goal of the first phase was to develop and administer a large pool of items from which overlength ASVAB forms could be constructed. The exception to this were the speeded subtests, Numerical Operations (NO) and Coding Speed (CS), which are not pretested in this initial tryout of the experimental items, but are administered in Phase II with just enough items to be operational length forms. The intent of the item writing part of Phase I was to produce sufficient items so that four ASVAB forms could be constructed. Two unique AFQT portions would be combined with two unique non-AFQT portions to provide a total of four parallel forms designated as ASVAB Forms 18a, 18b, 19a, and 19b.

The first step in this phase was to write items. Items were written by subject matter experts with the guidance that the same content domains as the

reference form be covered. A summarized taxonomy for the power subtests on the reference form, ASVAB Form 8a, and the percentage of items needed per content category are presented in Table 3. As a rule of thumb, at least 3 times as many items as were needed for the final forms were written. This allows for the discarding of those items not meeting the required standards. Table 4 presents the number of items written for each subtest so that 2 unique sets of AFQT and non-AFQT subtests could ultimately be constructed.

Each of the subject-matter experts was instructed on rules for proper item writing such as those found in Wesman (1971). In addition, these rules also included (a) avoiding the use of the item-options "all of the above" and "none of the above" and their variants, (b) arranging alternatives in ascending or descending order based on length (except in numerical items where they are arranged on magnitude), and (c) avoiding "clueing" the correct answer in any fashion. Further, items from existing and previous military tests and military enlistment and federal government qualification test study guides were not acceptable.

Table 3. ASVAB Subtest Taxonomy for Power Subtests

Subtest	Code	Areas (percentage of items)
GS	1	Life science (45%)
	2	Physical science (45%)
	3	Earth science (10%)
AR	1	Rearrangement of basic operations (35%)
	2	Rate/fraction problems (35%)
	3	Percentage problems (15%)
	4	Other (time, distance, area, etc.) (15%)
WK	1	Nouns (35%)
	2	Verbs (30%)
	3	Adjectives (35%)
PC	1	Literal detail (40%)
	2	Paraphrase/summarize (40%)
	3	Inferences/applications (20%)
AS		Auto
	1	Engines (21.5%)
	2	Body/drive train (21.5%)
	3	Electronics (7%)
		Shop
	4	Tools (35%)
	5	Materials (15%)
MK	1	Fractions/factoring (25%)
	2	Geometry (25%)
	3	Exponents/polynomials (15%)
	4	Equation solving (30%)
	5	Other (5%)
MC	1	Simple machines (10%)
	2	Basic compound machines (40%)
	3	Complex compound machines/structural components (20%)
	4	Mechanical concepts (30%)
EI	1	Theory & principles (20%)
	2	Circuit diagrams & wiring (10%)
	3	Power & electricity (40%)
	4	Tools & regulating devices (30%)

Table 4. Number of Items Required

Subtest	Number of forms	Number of required items per form	Total number of items required
GS	2	25	150
AR	2	30	180
WK	2	35	210
PC	2	15	90
AS	2	25	150
MK	2	25	150
MC	2	25	150
EI	2	20	120
TOTAL			1,200

Acceptable experimental items were assembled into 38 booklets that contained items from only one subtest. An item was subjectively determined to be acceptable if (a) it followed the good item writing practices described earlier, (b) covered one of the content areas listed in Table 3, and (c) was not determined by subjective examination to be offensive to subgroups of the population. Eight subtest booklets containing 8a items were also constructed. The 8a booklets were constructed so that they would be the same length as the experimental booklets. The 8a booklets were made overlength by adding experimental items after the 8a items. The number of items in the 8a and experimental booklets are presented in Table 5.

Table 5. Number of Items Used in Experimental Booklets

Subtest	Number of booklets	Number of items per booklet	
		8a	Experimental
GS - 8a	1	25	10
GS - Experimental	4	-	35
AR - 8a	1	30	-
AR - Experimental	6	-	30
WK - 8a	1	35	14
WK - Experimental	4	-	49
PC - 8a	1	15	6
PC - Experimental	4	-	21
AS - 8a	1	25	10
AS - Experimental	4	-	35
MK - 8a	1	25	10
MK - Experimental	4	-	35
MC - 8a	1	25	10
MC - Experimental	4	-	35
EI - 8a	1	20	15
EI - Experimental	4	-	35
TOTAL	42	200	1,235

Subjects

The experimental booklets were administered to 2,539 male and female basic trainees at Lackland AFB, Texas from October, 1984 through March, 1985.

Responses were recorded on machine scorable answer sheets. Each booklet was administered to approximately 250 examinees. A total of 200 reference and 1,235 experimental items were administered.

Data Analysis

After the machine scorable answer sheets were read, the correct item responses for each subject were summed to create subtest scores. Some data editing was performed in this phase prior to further analysis and included "cleaning up" the booklet number that was scanned from the answer sheet to ensure the application of the correct answer key to score the item responses.

For each power subtest, classical item statistics were computed for the total group and for male, female, white, hispanic and black subgroups, where sample sizes permitted. The classical item statistics included the difficulty levels of the items, defined as the percentage of examinees selecting the correct item option, and item discrimination, the biserial correlation between item and total test scores. Biserial correlations between distractor responses and total test scores were also calculated.

The goal of Phase I was to develop two unique overlength versions of the eight power subtests, each containing 15% to 20% more items than needed in the final operational subtest. Each of the overlength forms was designed to match Form 8a in content and item statistics as closely as possible. Table 6 shows the number of items in each overlength subtest.

Table 6. Number of Items Needed for Each Subtest

Subtest	Number of items
GS	35
AR	40
WK	45
PC	21
NO[a]	50
CS[a]	84
AS	35
MK	35
MC	35
EI	30

[a] *Speeded subtests are created in operational length in Phase II.*

The development of overlength subtests was accomplished by matching experimental items with the Form 8a items in terms of classical item statistics and content. Prior to matching the experimental items with the Form 8a items, the experimental items were examined for statistical acceptability. For an experimental item to be statistically acceptable, the item is required to have a difficulty value equal to or greater than .30 and a discrimination value equal to or greater than .35. These criteria were determined from examination of Form 8a statistical minimae. In addition, an experimental item was judged to be statistically acceptable if responses to distractors were not positively correlated with total test scores. It was possible that some of the reference form items would exhibit item statistics that were not within the desirable range as described for experimental items. In

that rare case an experimental item was deemed a match with the reference form item if the item was statistically close to the reference item but within acceptable criterion values.

Classical item statistics were used to match difficulty and discrimination of experimental form items with one another and with the reference form in developing overlength forms. The matching of experimental items to 8a items was a computer-aided process. If the experimental item's difficulty and discrimination values were within $\pm$ .05 of an 8a item's difficulty and discrimination values, the items were considered a match. To achieve close parallelism, highest priority was given to matching difficulties and numbers of illustrations, where applicable. Moderate priority was given to matching the taxonomic categories. Less priority was given to other factors, such as matching the discrimination values.

Results and Discussion

Experience has shown that approximately one-third of the experimental items usually meet acceptable statistical standards of quality and are eligible for further consideration. In this instance, approximately half of the items met minimum statistical standards. In Word Knowledge, Arithmetic Reasoning, and Mathematics Knowledge, approximately two-thirds of the items were statistically acceptable; however, many of these otherwise qualified items could not be used because they did not match an 8a item statistically. The requirement to match experimental items with 8a items was not met. It would have been advantageous if each ASVAB Form 8a item had two or more matching, eligibe items. In actuality, some had one, while others had no matching items. It was therefore necessary to obtain additional items. A second pool of items was developed and was administered to basic trainees at Lackland AFB, Texas from September 1984 through January 1985. As with the initial item pool, each item was administered to approximately 250 examinees.

The same item acceptability criteria were used. Acceptable items were evaluated to determine if their difficulty and discrimination values corresponded to the unmatched ASVAB Form 8a items.

With the supplemental items added to the pool, approximately 90% of the ASVAB Form 8a items had two or more matching items (one for each of the two new versions). For the remaining unmatched ASVAB Form 8a items, the nearest matches available were identified. Typically, the deviation between experimental and 8a item difficulty values fell within $\pm$ .05. However, the deviation between discrimination values was relaxed to approximately $\pm$ .15 in order to achieve matches.

After the two best matches for each ASVAB Form 8a item were identified, they were assigned to one of two alternate test forms. Assignment to forms was an iterative process. Adjustments were made to ensure that the forms were parallel with respect to mean difficulties, mean discriminations, taxonomic balance, and equal numbers of illustrated items.

After each version of the experimental test had exactly one item representing each ASVAB Form 8a item, the "extra" overlength items were identified. Selection of these items was primarily on the basis of the degree to which they enhanced the parallelism of the experimental forms. If an originally selected item had less than average similarity with its Form 8a counterpart, an extra item matching the same ASVAB Form 8a item was given special consideration for inclusion in the overlength

form. Also, illustrated items were given a higher priority for selection to ensure a sufficient number of them for the operational length forms.

After statistically parallel forms were assembled, their content was reviewed for internal irregularities such as items which were too similar to other items on the same test, thus clueing them. When such problems were discovered, these items were exchanged for statistically and taxonomically equivalent ones.

Paragraph Comprehension (PC) required special care. These items are not independent; several of them may refer to a common paragraph. With only 15 PC items in the operational length form, there is less freedom for fine-tuning this subtest through item selection. In addition, efforts were made to avoid repetitious paragraph content within a form and to control the lengths of the paragraphs so that overall paragraph length was comparable for the two new forms.

Frequencies of the keyed alternatives were counted after final item selections. When some alternatives were overrepresented, the alternative order was shuffled until alternative distribution was adequately balanced. For some of the subtests, the responses must be presented according to length (e.g., Word Knowledge) or in ascending or descending order (e.g., Arithmetic Reasoning) and therefore shuffling of alternatives was limited.

III. PHASE II - DEVELOPMENT OF OPERATIONAL LENGTH FORMS

The goal of Phase II was to develop operational length forms from the overlength forms developed in Phase I. Operational length Numerical Operations (NO) and Coding Speed (CS) subtests were developed at the beginning of this phase using the ASVAB 8a taxonomy as a guide. Since testing time at Recruit Training Centers (RTCs) is limited, the overlength power subtests developed in Phase I along with operational length speeded subtests developed in Phase II could not be administered as full batteries. The subtests were divided into three partial batteries and administered with like-named Form 8a subtests in a counter-balanced design to RTC recruits of all four services.

For each of the ten ASVAB subtests, there were two experimental versions, designated Version 1 and Version 2, plus the Form 8a version. To counterbalance the order of administration between the experimental versions and Form 8a, one set of booklets contained experimental Version 1 followed by Form 8a, while another set of booklets contained the same forms but with the experimental versions presented after Form 8a. This counterbalancing also was applied to experimental Version 2. Because administration time was limited to approximately 3 hours, it was necessary to construct partial booklets for each of three different subtest clusters - a total of 12 partial booklets (3 subtest clusters x 2 experimental versions x 2 experimental/reference test orders). One set of partial booklets contained Electronics Information (EI), Arithmetic Reasoning (AR), and Numerical Operations (NO) experimental and 8a subtests. The second set of partial booklets contained Auto and Shop Information (AS), Paragraph Comprehension (PC), Mechanical Comprehension (MC), and Coding Speed (CS) experimental and 8a subtests, while the third partial booklet contained General Science (GS), Word Knowledge (WK), and Mathematics Knowledge (MK).

Subjects

The 12 overlength booklets were tested at RTCs of all service to provide samples more representative of the ability of the overall service recruit population than that provided by the Air Force recruits used in Phase I testing. Testing was conducted from October 1986 through December 1986. Testing sites for each service were as follows:

Army - Ft. Jackson, South Carolina
Navy - Great Lakes, Illinois
Orlando, Florida
Air Force - Lackland AFB, Texas
Marine Corps - Parris Island, South Carolina.

Each booklet was administered to approximately 500 examinees. Given the counterbalanced design, each experimental items was in 2 of the 12 booklets, thus each experimental version was administered to approximately 1,000 examinees.

Data Analysis

Before computing item statistics, data editing procedures described in Ree, Mathews, Mullins, and Massey (1982) were applied to the data. First, the booklet number encoded by an examinee was verified. If the booklet number was coded incorrectly on the answer sheet or was missing, the answer keys for both forms of one subtest were applied to the answer sheet to determine the correct form identity. The easiest subtest in each booklet was chosen for this purpose. For the EI, AR, and NO partial booklet, the NO subtest was used for this purpose. For the AS, PC, MC, and CS partial booklets, the CS subtest was used to determine the booklet identity; and in the GS, WK, and MK partial booklets, WK was used. Cases were discarded if the booklet identity could not be established in this manner.

The next step in data editing involved scoring each subtest; the subtest scores were used to identify examinees scoring below chance level on the various subtests. Data for these individuals were discarded.

As in Phase I, classical item statistics were computed for each subtest. Because the ASVAB Form 8a included experimental items to make it agree in length with the experimental forms, item analyses for the Form 8a version were performed twice; once to obtain item statistics for the operational Form 8a and a second time to obtain statistics for the overlength Form 8a.

For the experimental test versions, the overlength subtest was scored and the items analyzed. The item statistics were examined for acceptability. If the difficulty or discrimination was less than .30 or .35, respectively, the item was deemed not acceptable for the operational length form. Also, if an item distractor had a positive biserial correlation with total test score, the item was deemed unacceptable. The item statistics were used to match difficulty values and discrimination values with the 8a items. As in Phase I, an experimental item was considered a match with a Form 8a item if the corresponding difficulty and discrimination values were within $\pm$ .05. Parallel forms for the ASVAB nonspeeded subtests were constructed using the ASVAB Form 8a content taxonomy, difficulties, and discriminations.

Finally, trial equatings of the operational length experimental subtests with the like-named 8a subtests were accomplished. The purpose of these equatings was to

determine whether the scores on the new subtests could be placed on the Form 8a score scale. Equipercentile and z-score (or linear) equatings were performed. Equipercentile equating was accomplished by obtaining the score distributions for the experimental and 8a subtests and defining scores that cut off the same percent of their respective distributions as equal. In addition, these equipercentile equatings were post-smoothed using linear, quadratic, and cubic polynomial regresion functions. Linear equating was accomplished by defining scores on the experimental and 8a subtests as equivalent if they had identical standard scores (z-scores) within their respective distributions. Differences among the linear, raw equipercentile, and post-smoothed equipercentile equatings were computed for comparison purposes. Bias, average absolute deviation (AAD), and root mean squared deviation (RMSD) indices were computed from the distributions of deviations.

Results and Discussion

Two parallel operational length high school forms were created by selecting items based on the all services RTC overlength testing. Items were retained on the basis of item statistic matches to the operational length ASVAB Form 8a so that the two operational length high school forms would be statistically parallel to each other and to the ASVAB Form 8a reference test. The best items were selected for each of the two test versions on the basis of simultaneous matching of their difficulties and discriminations to 8a items. The experimental tests also had to match the taxonomic representation of ASVAB Form 8a.

Items were culled if they did not statistically match a form 8a item, or if they were in a taxonomic area already fully represented. Items were also rejected if they were too similar to another item in the same version, clued another item's correct response, or were statistically flawed (the difficulty was less than .30, the discrimination was less than .35, or an item distractor had a positive biserial correlation value). The overlength items were treated as a pool from which any item could be used depending on how well it matched the characteristics of ASVAB Form 8a items as measured during the all service RTC testing.

Ideally, an item administered in a given overlength test version at the RTCs remains in that same version for the operational length test. However, to create operational length subtests that are statistically and taxonomically parallel to each other and to ASVAB Form 8a, it was necessary to switch several items among versions. This procedure was not necessary for any of the AFQT subtests (AR, WK, PC, or NO), but was necessary for four of the non-AFQT subtests. Two items were switched in Auto/Shop Information, four in Mechanical Comprehension, five in Electronics Information, and 12 in Mathematics Knowledge.

After the operational length tests were developed, the absolute and signed differences between their item difficulty and discrimination values and that of their corresponding ASVAB Form 8a item were calculated. The absolute and signed differences indicated the degree of variation from the ASVAB Form 8a reference test. Effort was exerted to minimize the number of selected items that exceeded an absolute difference of .05.

Tables 7 and 8 show the taxonomy representations, average difficulties, average discrimination values, and number of illustrated items for each subtest. As can be seen from these tables, the average difficulty of the 8a and experimental versions are within approximately .01 of each other. The discrimination indexes between forms are within approximately .05. For some of the subtests the taxonomic representation was identical across forms. For other subtests, the

Table 7. Percent of Items in Each Taxonomic Area and Average Difficulty and Discrimination Values for Operational Length AFQT Subtests[a]

Test	No. of items	Taxonomy area					Ill.	Average	
								Diff.	Disc.
AR		1	2	3	4				
8a	30	.35	.35	.15	.15			.624	.548
V1	30	.20	.33	.17	.30			.635	.515
V2	30	.23	.33	.13	.30			.633	.524
WK		1	2	3	Inc.	Comp.			
8a	35	.35	.30	.35	.60	.40		.799	.581
V1	35	.23	.37	.40	.63	.37		.797	.591
V2	35	.20	.43	.37	.63	.37		.808	.593
PC		1	2	3					
8a	15	.40	.40	.20				.775	.578
V1	15	.40	.40	.20				.782	.598
V2	15	.40	.40	.20				.789	.581
MK		1	2	3	4	5			
8a	25	.25	.25	.15	.30	.05	3	.597	.545
V1	25	.32	.24	.16	.20	.08	3	.593	.584
V2	25	.24	.32	.12	.24	.08	3	.599	.592

[a] *NOTES 1. Taxonomy areas are identifed in the Appendix.*
2. In WK, Inc. denotes that the item stem is not a complete sentence but is in the form "[word] most nearly means...." Comp. denotes that the item stem is a complete sentence.
3. Ill. is an abbreviation for number of items illustrated.
4. Diff. = Difficulty.
Disc. = Discrimination.

Table 8. Percent of Items in Each Taxonomic Area and Average Difficulty and Discrimination Values for Operational Length Non-AFQT subtests[a]

Test	No. of items	Taxonomy area							Ill.	Average	
										Diff.	Disc.
GS		1	2	3							
8a	25	.48	.48	.04						.597	.545
V1	25	.52	.40	.08						.593	.584
V2	25	.44	.48	.08						.599	.592
AS		1	2	3	4	5	Auto	Shop			
8a	25	.22	.22	.07	.35	.15	.56	.44	5	.668	.548
V1	25	.24	.20	.12	.32	.12	.56	.44	6	.675	.584
V2	25	.24	.16	.12	.36	.12	.52	.48	5	.659	.578
MC		1	2	3	4						
8a	25	.10	.40	.20	.30					.648	.506
V1	25	.12	.48	.08	.32					.660	.531
V2	25	.16	.44	.08	.32					.652	.523
EI		1	2	3	4						
8a	20	.20	.10	.40	.30				1	.639	.506
V1	20	.20	.10	.40	.30				2	.630	.491
V2	20	.20	.10	.40	.30				2	.629	.500

[a] *NOTES 1. Taxonomy areas are identifed in the Appendix.*
2. Ill. is an abbreviation for number of items illustrated.
3. Diff. = Difficulty.
Disc. = Discrimination.

taxonomic differences between forms are relatively small, especially given difficulty and discrimination matching requirements.

Data from the final subtest configuration were available for some of the subtests (those subtests that did not switch items between Versions 1 and 2), and are shown in Table 9 for Versions 1 and 2. Comparable data were not available for other subtests because items from one overlength form were exchanged with items from the other overlength form in the selection process and therefore total subtest scores could not be computed. Table 9 shows that the test statistics for Versions 1 and 2 are generally equal.

Table 9. Subtest Descriptive Statistics for Operational Length Forms

	General Science		Arithmetic Reasoning	
	Version 1	Version 2	Version 1	Version 2
No. Items	25	25	30	30
Mean	17.077	16.817	19.039	18.987
Median	17	17	18	19
Variance	17.093	15.251	27.766	27.715
SD	4.134	3.905	5.269	5.265
Skew	-0.222	-0.290	0.150	0.027
Kurtosis	-0.655	-0.371	-0.698	-0.759
Minimum	6	5	7	7
Maximum	25	25	30	30
KR-20	0.761	0.737	0.813	0.815
SEM	2.020	2.003	2.279	2.264
No. Examinees	1045	1024	964	987

	Word Knowledge		Paragraph Comprehension	
	Version 1	Version 2	Version 1	Version 2
No. Items	35	35	15	15
Mean	27.906	28.271	11.732	11.839
Median	29	29	12	12
Variance	23.850	20.992	6.722	5.920
SD	4.884	4.582	2.593	2 433
Skew	-0.683	-0.767	-0.777	-0.856
Kurtosis	0.178	0.332	-0.011	0.407
Minimum	9	11	3	3
Maximum	35	35	15	15
KR-20	0.829	0.810	0.695	0.660
SEM	2.019	1.998	1.432	1.419
No. Examinees	1021	1032	978	988

Trial equatings of recommended operational length subtests. Trial equatings were conducted to determine if scores on the two versions of the experimental subtests corresponded to each other and to the scores on like-named Form 8a subtests.

Equipercentile trial equatings using raw score frequency distributions are desirable on the 10 ASVAB subtests. However, not all of these equatings could be accomplished. To improve the parallelism of some of the subtests, items originally appearing on experimental Version 1 were switched to experimental Version 2 and vice versa. This item swapping occurred in the AS, MK, MC, and EI subtests. Therefore, these subtests could not be equated because the same individuals did not

take all of the items in a final form.

Equipercentile (with and without smoothing) and linear trial equatings were performed on the edited raw score frequency distributions for both experimental versions of GS, AR, WK, PC, NO, and CS. Tables in the Appendix show the results of these trial equatings. Because these are trial equatings of untested operational length forms, a choice of the preferred equating and smoothing was not warranted at this time. However, examination of the equating resuits indicated that for each experimental subtest a particular equating could be selected that would closely replicate the first four moments of the distribution of the like-named reference subtest.

IV. SUMMARY

Two unique sets of ASVAB subtests that are combined to yield four new parallel forms of the ASVAB- Forms 18a, 18b, 19a, and 19b- for use in the Student Testing Program were developed. These four forms are constructed to be parallel with one another and to the reference test, Form 8a, in terms of difficulty, discrimination, and taxonomy.

REFERENCES

Department of Defense. (1984). *Technical supplement to the counselor's manual for ASVAB-14*. North Chicago, IL: U. S. Military Entrance Processing Command.

Office of the Assistant Secreatry of Defense. (1982). *Profile of the American youth: Field report*. Washington, DC: Author.

Office of Naval Research. (1985). *User's manual for the MicroCAT testing system*. Arlington, VA: Author.

Ree, M. J., Mathews, J. J., Mullins, C. J., & Massey, R. H. (1982). *Calibration of Armed Services Vocational Aptitude Battery Forms 8, 9, and 10* (AFHRL-TR-81-49, AD-A114 714). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Ree, M. J., Mullins, C. J., Mathews, J. J., & Massey, R. H. (1982). *Armed Services Vocational Aptitude Battery: Item and factor analyses of forms 8, 9, and 10* (AFHRL-TR-81-55, AD-A113 465). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Wesman, A. G. (1971). Writing the test item. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council of Education.

APPENDIX A: Equating Results

Table A-1. Results of Equating GS Version 1 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	1045	1045
Minimum value	6.0000	7.0000
Maximum value	25.0000	25.0000
Mean	17.0766	16.7502
Standard deviation	4.1363	3.9185
Skew	-.2221	-.1388
Kurtosis	2.3493	2.3507

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	-.0526	.1499	.1766
Linear smooth		.0526	.0585
Quadratic smooth		.0948	.1240
Cubic smooth		.0981	.1334

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.1421	.1667
Quadratic smooth	.1036	.1257
Cubic smooth	.0926	.1157

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	16.7546	3.9401	-.1370	2.3651
Linear smooth	16.8098	3.9367	-.2221	2.3493
Quadratic smooth	16.7594	3.9715	-.1603	2.3086
Cubic smooth	16.7511	3.9495	-.1449	2.3452
Z-score	16.7502	3.9184	-.2221	2.3493

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-2. Results of Equating GS Version 2 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	1024	1024
Minimum value	5.0000	7.0000
Maximum value	25.0000	25.0000
Mean	16.8174	17.0215
Standard deviation	3.9071	3.8350
Skew	-.2901	-.0663
Kurtosis	2.6346	2.4287

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	-.2355	.3446	.4762
Linear smooth		.2518	.3102
Quadratic smooth		.3197	.4547
Cubic smooth		.3017	.4593

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.3244	.3608
Quadratic smooth	.1145	.1510
Cubic smooth	.1056	.1339

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	17.0185	3.8571	-.0727	2.4366
Linear smooth	17.1961	3.7042	-.2901	2.6346
Quadratic smooth	17.0088	3.7944	-.1237	2.4701
Cubic smooth	17.0294	3.8315	-.1444	2.4261
Z-score	17.0210	3.8339	-.2917	2.6324

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-3. Results of Equating AR Version 1 and 8a

Test descriptive statistics

	Version 1	8a
Number of observations	964	964
Minimum value	7.0000	8.0000
Maximum value	30.0000	30.0000
Mean	19.0394	18.6152
Standard deviation	5.2721	5.5385
Skew	.1500	.1997
Kurtosis	2.3047	2.1226

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	-.1365	.2951	.4343
Linear smooth		.1993	.2443
Quadratic smooth		.2278	.3286
Cubic smooth		.2509	.4093

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.2868	.3569
Quadratic smooth	.2569	.2891
Cubic smooth	.1176	.1356

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	18.6135	5.5461	.1942	2.1279
Linear smooth	18.7356	5.3813	.1500	2.3047
Quadratic smooth	18.6314	5.4169	.2342	2.3002
Cubic smooth	18.6180	5.5376	.2016	2.1549
Z-score	18.6125	5.5330	.1457	2.2970

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-4. Results of Equating AR   ion 2 and 8a

Test descriptive statistics

	Version 2	8a
Number of observations	987	987
Minimum value	7.0000	8.0000
Maximum value	30.0000	30.0000
Mean	18.9868	18.8278
Standard deviation	5.2672	5.7051
Skew	.0270	.1669
Kurtosis	2.2439	2.0442

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	-.1668	.4997	.6346
Linear smooth		.2687	.3236
Quadratic smooth		.3272	.4710
Cubic smooth		.3545	.5592

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.4830	.5519
Quadratic smooth	.3745	.4722
Cubic smooth	.2346	.2893

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	18.8240	5.7179	.1605	2.0537
Linear smooth	18.9666	5.4842	.0108	2.2163
Quadratic smooth	18.8223	5.5037	.1165	2.2055
Cubic smooth	18.8333	5.7151	.0825	2.0178
Z-score	18.8132	5.6755	.0031	2.2040

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-5. Results of Equating WK Version 1 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	1021	1021
Minimum value	9.0000	11.0000
Maximum value	35.0000	35.0000
Mean	27.9100	27.8648
Standard deviation	4.8873	4.7095
Skew	-.7038	-.7469
Kurtosis	3.2583	3.3949

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	.0692	.3291	.4417
Linear smooth		.0692	.0715
Quadratic smooth		.1519	.1701
Cubic smooth		.2729	.3277

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.3270	.4359
Quadratic smooth	.3408	.4077
Cubic smooth	.2649	.2962

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	27.8613	4.7214	-.7446	3.3840
Linear smooth	27.8091	4.7207	-.7038	3.2583
Quadratic smooth	27.8038	4.8384	-.6306	3.0619
Cubic smooth	27.8772	4.7954	-.7792	3.1853
Z-score	27.8648	4.7095	-.7038	3.2583

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-6. Results of Equating WK Version 2 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	1032	1032
Minimum value	11.0000	9.0000
Maximum value	35.0000	35.0000
Mean	28.2713	28.0446
Standard deviation	4.5840	4.6267
Skew	-.7684	-.7130
Kurtosis	3.3373	3.3362

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	.0816	.3467	.5526
Linear smooth		.1667	.1977
Quadratic smooth		.2281	.3074
Cubic smooth		.3022	.4431

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.3553	.5160
Quadratic smooth	.3627	.4592
Cubic smooth	.2519	.3301

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	28.0382	4.6338	-.7098	3.3174
Linear smooth	28.0920	4.7374	-.7722	3.3394
Quadratic smooth	28.1110	4.5799	-.8923	3.6743
Cubic smooth	28.0030	4.5849	-.7275	3.5301
Z-score	28.0445	4.6267	-.7684	3.3373

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-7. Results of Equating PC Version 1 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	978	978
Minimum value	3.0000	4.0000
Maximum value	15.0000	15.0000
Mean	11.7321	11.5501
Standard deviation	2.5940	2.3765
Skew	-.7778	-.8510
Kurtosis	2.9987	3.2515

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	.0652	.1253	.1592
Linear smooth		.1005	.1210
Quadratic smooth		.0898	.1265
Cubic smooth		.1051	.1313

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.0878	.1035
Quadratic smooth	.0809	.0967
Cubic smooth	.0776	.0901

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	11.5536	2.4029	-.8413	3.2430
Linear smooth	11.5593	2.4472	-.7778	2.9987
Quadratic smooth	11.5587	2.4206	-.8126	3.0855
Cubic smooth	11.5643	2.4188	-.8470	3.1160
Z-score	11.5501	2.3765	-.7778	2.9987

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-8. Results of Equating PC Version 2 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	988	988
Minimum value	3.0000	4.0000
Maximum value	15.0000	15.0000
Mean	11.8390	11.7004
Standard deviation	2.4344	2.3579
Skew	-.8571	-.7797
Kurtosis	3.4197	3.1993

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	-.0216	.0832	.1015
Linear smooth		.0336	.0404
Quadratic smooth		.0485	.0655
Cubic smooth		.0660	.0750

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.0735	.0931
Quadratic smooth	.0647	.0775
Cubic smooth	.0541	.0683

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	11.6941	2.3755	-.7813	3.2111
Linear smooth	11.6961	2.3357	-.8571	3.4197
Quadratic smooth	11.6960	2.3722	-.8051	3.2867
Cubic smooth	11.6868	2.3765	-.7610	3.2350
Z-score	11.7004	2.3579	-.8571	3.4197

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-9. Results of Equating NO Version 1 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	999	999
Minimum value	11.0000	14.0000
Maximum value	50.0000	50.0000
Mean	41.1952	42.9740
Standard deviation	7.6918	7.0956
Skew	-.9572	-1.2269
Kurtosis	3.3029	4.0585

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	.6105	.9505	1.3489
Linear smooth		.9049	1.0971
Quadratic smooth		.7822	1.1725
Cubic smooth		.9699	1.2192

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.6747	.8366
Quadratic smooth	.4595	.6829
Cubic smooth	.3934	.5458

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	42.9586	7.0999	-1.2265	4.0630
Linear smooth	42.9252	7.4266	-1.0604	3.4771
Quadratic smooth	42.9217	7.1566	-1.1395	3.8356
Cubic smooth	42.9586	7.1002	-1.2664	4.1183
Z-score	42.8761	6.9908	-1.0073	3.3785

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-10. Results of Equating NO Version 2 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	980	980
Minimum value	11.0000	8.0000
Maximum value	50.0000	50.0000
Mean	41.6459	42.9591
Standard deviation	7.6682	7.4531
Skew	-.9719	-1.3277
Kurtosis	3.3499	4.4832

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	.8033	1.3081	1.6870
Linear smooth		1.2230	1.4774
Quadratic smooth		1.0679	1.6081
Cubic smooth		1.2120	1.6251

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.6887	.8744
Quadratic smooth	.4485	.5259
Cubic smooth	.2937	.4252

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	42.9436	7.4521	-1.3249	4.4776
Linear smooth	42.9394	7.8873	-1.1035	3.6108
Quadratic smooth	42.9326	7.5044	-1.2129	4.1745
Cubic smooth	42.9441	7.4387	-1.3000	4.3934
Z-score	42.8365	7.3268	-1.0262	3.4430

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-11. Results of Equating CS Version 1 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	1003	1003
Minimum value	1.0000	1.0000
Maximum value	84.0000	84.0000
Mean	55.0668	56.2453
Standard deviation	15.7579	15.4080
Skew	-.3030	-.3242
Kurtosis	3.1509	3.1348

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	-.0048	.7281	1.0391
Linear smooth		.1069	.1594
Quadratic smooth		.2157	.2605
Cubic smooth		.2075	.2844

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.7273	1.0331
Quadratic smooth	.7381	1.0073
Cubic smooth	.7373	1.0031

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	56.2430	15.4075	-.3244	3.1335
Linear smooth	56.2841	15.4362	-.3164	3.1468
Quadratic smooth	56.3778	15.3198	-.3526	3.2132
Cubic smooth	56.3460	15.3107	-.3389	3.2217
Z-score	56.2271	15.3747	-.3134	3.1476

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*

Table A-12. Results of Equating CS Version 2 and 8a

Descriptive statistics of the tests

	Version 1	8a
Number of observations	997	997
Minimum value	10.0000	10.0000
Maximum value	84.0000	84.0000
Mean	55.3160	56.7503
Standard deviation	14.2504	14.9307
Skew	-.1627	-.1729
Kurtosis	2.8409	2.8615

Deviations from z-score linear equating

Equipercentile equating method	Bias	AAD	RMSD
Raw	.0590	.8458	1.0480
Linear smooth		.2949	.5224
Quadratic smooth		.3127	.5046
Cubic smooth		.3950	.4874

Deviations from raw scores (measure of fit)[a]

Equipercentile equating method	AAD	RMSD
Linear smooth	.8105	1.0167
Quadratic smooth	.8135	.9984
Cubic smooth	.7928	.9695

Reproduced moments of the distribution from equating transformations

Equating	Mean	Standard deviation	Skew	Kurtosis
Raw equipercentile	56.7448	14.9266	-.1738	2.8586
Linear smooth	56.6955	14.9202	-.2187	2.7972
Quadratic smooth	56.8011	14.8574	-.2438	2.8365
Cubic smooth	56.8767	14.9507	-.2641	2.7996
Z-score	56.6714	14.7797	-.2109	2.8020

[a] *Bias is omitted because regression smoothing produces zero biased estimates.*